Northwestern University

2001 Sheridan Road 580 Leverone Hall Evanston, IL 60208-2014 USA

Discussion Paper #1091

June, 1994

"A Chaotic Exploration of Aggregation Paradoxes"

Donald G. Saari
Northwestern University

www.kellogg.nwu.edu/research/math

CMS-EMS
The Center for
Mathematical Studies
in Economics &
Management Sciences

A CHAOTIC EXPLORATION OF AGGREGATION PARADOXES

Donald G. Saari

Department of Mathematics, Northwestern University

Abstract. Paradoxes from statistics and decision sciences form amusing, yet intriguing mathematical puzzles. On deeper examination, they constitute serious problems that could cause us, unintentionally, to adopt inferior alternatives. It is indicated here how ideas from "dynamical chaos" and orbits of symmetry groups can be modified and combined to create a mathematical theory to understand, classify, and find new properties of these puzzling phenomena.

Paradoxes are intriguing. By a paradox, I mean a mathematically counterintuitive conclusion. To illustrate, the likelihood of "Heads" with a flipped penny is approximately $\frac{1}{2}$. Presumably, the same answer holds for a penny spinning on edge, but, instead, it is only about 0.30. This is a surprise, not a paradox because an examination discloses that the slightly heavier "Head" tilts the axis of rotation. A paradox requires a more subtle mathematical structure.

Suppose 15 friends decide to choose a common beverage [S7]. Their preferences

- six have "milk $\succ$ wine $\succ$ beer,"
- five have "beer $\succ$ wine $\succ$ milk," and
- four have "wine $\succ$ beer $\succ$ milk,"

define the plurality ranking "milk $\succ$ beer $\succ$ wine" with the vote $6:5:4$. If bottom-ranked wine is not available, we expect nothing to change; milk should remain the top-choice. However, 60% of this group prefers "beer $\succ$ milk" with the vote of $9:6$! They even prefer "wine $\succ$ milk" and "wine $\succ$ beer" with similarly decisive votes. These pairwise comparisons, then, suggest that the plurality ranking completely reverses what the voters really want! Instead of being the beverage of choice, these voters view milk as their inferior alternative. This is a paradox.

A probability example can be created by replacing the usual markings on three fair dice with the numbers (from a magic square)

$$
\begin{array}{cccc}
A & \boxed{8} & \boxed{1} & \boxed{6} \\
B & \boxed{3} & \boxed{5} & \boxed{7} \\
C & \boxed{4} & \boxed{9} & \boxed{2}
\end{array}
\tag{1.1}
$$

1991 *Mathematics Subject Classification.* 90A08, 62J15, 58F08.

Key words and phrases. Positional voting, voting, paradoxes, chaos, Kruskal Wallis test.

This essay is based on invited lectures for the AMS, South Bend, Indiana, March, 1991, and the MAA, Wisconsin section, April, 1992, and several colloquia. The research described in this paper is supported by NSF IST 9103180

where each number appears twice. Each person chooses and rolls a die where high score wins. For two dice, the sample space has nine pairs, so a die with the larger value in at least five of them is the better choice. As a count proves that $A \succ B$, $B \succ C$, it is reasonable to anticipate that $A \succ C$: instead, $C \succ A$.

Among the many statistical paradoxes, suppose a new cold remedy is tested in Evanston and in Recife where in both sites the data supports the new remedy. For figures, suppose out of the 144 subjects from Evanston who opt for the experimental approach, 54 regain health. From the 36 using the standard approach, only 12 regain health. As $\frac{54}{144} = \frac{3}{8} > \frac{12}{36} = \frac{1}{3}$, the experimental drug is better. In Recife, of the 36 who opted for the experimental approach, half regain health. Of the 144 in the control group, 66 regain health. As $\frac{1}{2} > \frac{66}{144}$, the experimental drug dominates again. In both locales, the experimental drug proved better than the standard approach, so, presumably, it should be chosen. In the aggregate with 180 subjects in each group, however, only 72 from the experimental group regained health compared to the 78 from the control group. This *Simpson's Paradox* (e.g., see [B, GM, HS]) demonstrates that the aggregated data can reverse the conclusion!

Paradoxes are amusing, but when recast in terms of an election for the departmental chair, the selection of the candidate for the sole tenure track position, a Presidential Candidate, the choice of economic or political decisions, or comparing medical tests it becomes evident that these mathematical puzzles have meaningful consequences that must be addressed. Indeed, with the vast numbers of daily decisions made in various contexts, one must expect that many of the mathematical paradoxes from the decision and statistical sciences have been manifested by groups unknowingly selecting inferior alternatives. This, in itself, underscores the critical importance of understanding these counterintuitive surprises.

These are well-studied issues, but progress has been severely limited. The reason is clear: a "paradox" is counterintuitive, so what does one look for? This is why, even with the huge literature for each of these fields (e.g., for voting, see the 72 page bibliography [K] or the surveys [Br, K2, Nu, O, St]), only relatively few paradoxes and properties of ranking systems have been found. The problem is further complicated by the assertion ([BO, BTT]) that it can be NP-hard to determine whether an election outcome can change when a non-winner drops out. In this article, I outline a new approach based on "chaotic dynamics" and symmetry groups that overcomes these severe limitations by extracting all possible ranking paradoxes and properties that can occur.

2. A Chaotic State of Affairs

The chaotic examination of aggregation procedures is based on the mathematical structures of "chaos" as outlined next. It is important to stress that conclusions such as period-doubling, homoclinic points, strange attractors, etc., are not and cannot occur in an analysis of paradoxes. These results intimately depend upon particular structures of dynamical systems that are missing here. Instead, what we borrow from dynamical systems is the change of perspective from a local to a global emphasis along with the mathematical insight motivating the development of "symbolic dynamics." Once modified, these conceptual ideas define new ways to analyze the paradoxes of aggregation procedures. Moreover, this "borrowing" from dynamics identifies a research program of parallel issues for discrete and "static" problems.

A way to review basic ideas from dynamics is to consider Newton's method [SU] for finding a zero of a polynomial $f(x)$. (For a general introduction, see [D].) If the initial guess, x_1, fails because $f(x_1) \neq 0$, then f is replaced by its linear approximation, $f(x_1) + (x - x_1)f'(x_1)$: the zero of this approximation, x_2, is the next iterate. As such, Newton's method has the standard geometric representation of Fig. 1. Of course, the iteration procedure fails at a critical point of f because the horizontal linear approximation has no zeros.

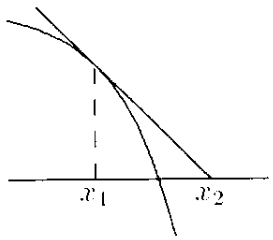

Figure 1. Newton's method.

Suppose we are to find the roots of the fifth-degree polynomial depicted in Fig. 2 where the critical points of f define the endpoints of the three labeled intervals. We know that if x_1 is in an unbounded interval, the iteration converges to the zero of that interval. The remaining challenge, then, is to understand what happens should all iterates of an initial point remain in $a \cup b \cup c$. Some iterates converge to a zero of f in this region, but what else can happen?

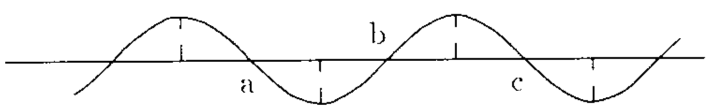

Figure 2. Regions for a fifth degree polynomial.

One way to discover "nonconvergent" properties of Newton's method is to examine the behavior of the orbit $x_1, x_2, \ldots, x_n, \ldots$ by experimenting with the choice of x_1. For instance, if x_{61} ends up near x_1, it is reasonable to expect from continuity considerations that x_1 can be modified to force $x_1 = x_{61}$: this means we should expect a period-sixty orbit to exist. This is a simplified, yet, not inaccurate description of how various properties were obtained. This approach, with its emphasis on local properties, can be technically difficult. Yet, by being property-specific, it doesn't tell us what else might happen. For instance, we must anticipate certain periodic orbits to exist; e.g., period-two orbits are easy to construct. Are there any other kinds? Can anything other than periodic orbits happen?

As an alternative way to address these issues, after $\{x_1, x_2, x_3, \ldots\}$ is determined, replace each iterate with a label identifying the interval in which it lives. For example, if $x_1 \in a, x_2 \in b, x_3 \in c, x_4 \in b, \ldots$, then the initial condition x_1 and its orbit defines the sequence $g_f(x_1) = (a, b, c, b, \ldots)$. As this listing of symbols is not random (it is ordered by the dynamics), call it a *word*. So, if $U^3 = \{a, b, c\}^N$ is the *universal set* consisting of all possible sequences where each entry is one of the three symbols, then Newton's method defines a mapping $g_f : a \cup b \cup c \to U^3$ where each initial iterate is identified with its word – an element of U^3. A word, then, specifies one kind of outcome. One way to measure the complexity of Newton's method is to find all possible words; that is, to find all of the entries in the *dictionary*

$$\mathcal{D}_f = \{g_f(x_1) \in U^3\} \subset U^3. \tag{2.1}$$

Common sense dictates that should $\mathcal{D}_f$ be a large subset of $\mathcal{U}^3$, then Newton's method admits a rich supply of complex, chaotic dynamics. On the other hand, with a limited number of words, not much can happen, so if $\mathcal{D}_f$ is a small subset of U^3, a reasonably benign dynamic must be anticipated. Thus the size of $\mathcal{D}_f$ serves as a crude complexity measure of the process. A way to understand Newton's method, then, is to characterize its dictionary $\mathcal{D}_f$. Notice the change in emphasis: instead of detecting particular features of Newton's method, the ambitious new goal is to completely catalogue all long term dynamical properties (i.e., all sequences in $\mathcal{D}_f$). This change in emphasis favoring global over local properties is key for our analysis and classification of aggregation paradoxes.

As developed in [SU], Newton's method is as complex and chaotic as possible with this dictionary measure because

$$\mathcal{D}_f = U^3. \tag{2.2}$$

Namely, choose any sequence generated by the letters a, b, c – an entry can even be determined by rolling a die – and there is an initial iterate whereby the jth iterate lands in the interval specified by the jth entry of the sequence. Therefore, for the sequence $s = (b, a, c, a, c, b, b, \ldots)$, Eq. 2.2 asserts there exists $x_1 \in b$ so that $x_2 \in a, x_3 \in c, x_4 \in a, x_5 \in c, \ldots$. With only slight extra effort, it follows that periodic orbits of any period following any pattern through the three intervals exist! Moreover, there also are orbits that bounce forever among the intervals in counterintuitive, non-periodic ways.

Particularly important for the analysis of paradoxes is that, because this technique characterizes all orbits, it answers the puzzling problem how to discover counterintuitive properties. So, by extending this dictionary approach to aggregation procedures, it becomes realistic to search for "everything that can happen;" namely, we can hope to discover all possible ranking paradoxes and properties. With this objective in mind, we need to understand how to find the entries of $\mathcal{D}_f$.

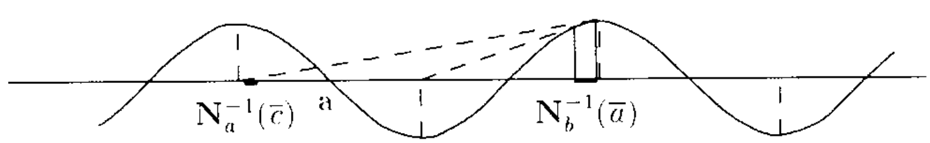

Figure 3. The iterated inverse image.

The method of proof (to establish that g_f is surjective) uses an "iterated inverse image" approach that I illustrate with the sequence $s = (b, a, c, a, c, b, b, \ldots)$. The goal is to keep refining the set of initial iterates that accomplishes each portion of the designated future. To see how to do this, if N_k is Newton's map restricted to interval $k = a, b, c$, then the set of iterates starting in b and ending in a is $N_b^{-1}(\overline{a})$ where $\overline{a}$ is the closure of the interval a. The key fact is that N_k maps an interval k onto $(-\infty, \infty)$. To see why, notice that as x moves closer to the left-hand endpoint of interval k, the straight line approximation of f becomes horizontal forcing its zero (the next iterate $N_k(x)$) far to the right. Similarly, for values of $x \in k$ near the right-hand endpoint, the line again approaches an horizontal position, but now it crosses the x-axis far to the left. Because $N_k : k \to (-\infty, \infty)$ is surjective, $N_b(\overline{a})$ is a nonempty closed subset of b. Geometrically, $N_b(\overline{a})$ is easy to determine as depicted in Fig. 3.

$N_b^{-1}(\overline{a})$ are the points in b that are mapped to $\overline{a}$. We don't care about reaching all points in a; only those for which the next iterate is in c. Using the same argument, we only want to reach $N_a^{-1}(\overline{c}) \subset a$. Thus, we want to find the subset of $N_b^{-1}(\overline{a})$ where its next iterate is in the much smaller target region $N_a^{-1}(\overline{c})$ (so that iterates start in b, go to a and then to c): clearly, this refinement is

$$b \supset N_b^{-1}(\overline{a}) \supset N_b^{-1}(N_a^{-1}(\overline{c})).$$

The idea now is obvious. To find all initial points satisfying the designated future, continue this iterated inverse image approach to obtain the nested sequence of nonempty, closed, bounded sets

$$(2.3) \quad b \supset N_b^{-1}(\overline{a}) \supset N_b^{-1}(N_a^{-1}(\overline{c})) \supset \cdots \supset N_b^{-1}(N_a^{-1}(\ldots(N_k^{-1}(\overline{k+1})\ldots)) \supset \ldots$$

A point in the intersection of all sets serves as an initial iterate satisfying the conditions of the sequence. As standard results from elementary analysis ensure that this intersection is nonempty, such an initial iterate exists.

Intuitively, the source of certain assertions, such as "sensitivity to initial conditions" and "Cantor set" constructions now become clear. Observe from Fig. 3 that the expanding nature of N_b requires $N_b^{-1}(\overline{a})$ to be a small set and $N_b^{-1}(N_a^{-1}(\overline{c}))$ to be much smaller. Yet, this small set $N_b^{-1}(N_a^{-1}(\overline{c}))$ contains all b points that pass through a and then are mapped onto c. As such, $N_b^{-1}(N_a^{-1}(\overline{c}))$ contains a distinct subset for each of the uncountable ways $(b, a, c, \ldots)$ can be continued. Consequently, most points of $N_b^{-1}(N_a^{-1}(\overline{c}))$ must be near regions with radically different orbits, so a slightly varied initial point could change regions and embark on a dramatically different future. This is the source of the sensitivity. Similarly, for each extension of $(b, a, c, \ldots)$, each step of the iterated inverse image approach identifies all points that eventually are mapped onto the next specified interval. Among these points is an open set ensuring convergence to the zero of that interval. Thus, to construct the set of nonconvergent points, an open set is excised at each step just as in the construction of the standard middle-thirds Cantor set.

3. Return to the paradoxes

In voting theory and statistics, paradoxes and troublesome properties of procedures typically are discovered by finding specific examples of voters' preferences or data. For instance, one of the better known voting paradoxes, designed by P. Fishburn [NR], illustrates the interesting behavior where the voters' sincere plurality election ranking of $A \succ B \succ C \succ D$ is reversed to $C \succ B \succ A$ when D, the bottom-ranked candidate, withdraws. Part of the difficulty of finding examples of this type is to even suspect that they occur. Then the method of proof finding a supporting example of voters' preferences often leads to combinatoric difficulties. Indeed, as argued in [BO, BTT], the computational difficulty to recognize that these difficulties can occur can be NP-hard. (But, "recognizing" that we are experiencing a difficulty and the likelihood it can occur are very different issues. As shown in [S7], paradoxes are surprisingly probable.)

Observe the close parallels between this traditional way to analyze voting and analyzing Newton's method guided by properties of particular initial iterates. Both approaches concentrate on local properties, so it is difficult, if not impossible, to even suspect what else could occur. (For instance, with the election reversal example, what could happen if C also decides to withdraw; or, if D returns but now A withdraws?) Then, even with limited results, the supporting proofs can be technically difficult. To remove these obstacles, one might hope to mimic the dynamical systems approach (applied above to Newton's method) to transform these decision problems into a mathematical framework emphasizing global properties. I introduce the ideas with the beverage example.

In dynamics, the initial condition x_1 is the "starting point." Voting, on the other hand, begins with a listing of each voter's preferences, a *profile*. This means that the listing of voter's preferences in the beverage example defines a particular 15-voter profile. Thus, *a profile becomes the "initial condition" for voting*; there are no restrictions on the number of voters.

Key to dynamics is the ordering of outcomes as imposed by "time;" in the Newton sequence this ordering is indicated by the subscript of the jth term, x_j. A natural "time" variable doesn't exist for voting; instead the goal is to compare election outcomes over different subsets of candidates. But, we can invent one. For instance, with $n = 3$ candidates $\{c_1, c_2, c_3\}$, the four subsets with two or more candidates are $S_1 = \{c_1, c_2\}$, $S_2 = \{c_1, c_3\}$, $S_3 = \{c_2, c_3\}$, $S_4 = \{c_1, c_2, c_3\}$. In general with $n \geq 3$ candidates, list the $2^n - (n+1)$ subsets of two or more candidates (in some fashion) as $S_1, S_2, \dots, S_{2^n-(n+1)}$. Replacing the ordering role of "*time*" from dynamics, then, is the integer indicating which subset of candidates is being considered.

The Newton method example emphasized the three subintervals labeled with the symbols a, b, c where the precise value of the jth iterate, x_j, is replaced with the cruder information identifying the subinterval in which it lives. In voting, the subscript j represents the subset of candidates S_j. Corresponding to the precise value of an iterate is the precise election tally of the S_j candidates; replace this tally with the cruder information specifying the election ranking. Thus, replacing the symbols $\{a, b, c\}$ for Newton's method is $\mathcal{R}(S_j)$ the set of all possible election rankings of S_j. For instance, $S_3 = \{c_2, c_3\}$ has the three symbols $\mathcal{R}(S_3) = \{c_2 \succ c_3, c_2 \sim c_3, c_3 \succ c_2\}$ indicating that either one candidate beats the other, or they tie. Similarly, $\mathcal{R}(S_4)$ has thirteen entries; six correspond to the six ways candidates can be ranked without ties, six are where there is one tie, and one corresponds to a complete tie.

In Newton's method, the Universal set consists of all possible sequences that can be constructed. Each entry is one of the three choices, so

$$U = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \cdots = \begin{pmatrix} a \\ b \\ c \end{pmatrix}^{N}.$$

In voting, the jth entry is the symbol from $\mathcal{R}(S_j)$ designating the ranking of the candidates from S_j. Therefore, the universal set is

$$\mathcal{U}^3 = \mathcal{R}(S_1) \times \mathcal{R}(S_2) \times \mathcal{R}(S_3) \times \mathcal{R}(S_4).$$

In the general case of n candidates, where there are $2^n - (n+1)$ subsets of two or more candidates, the universal set is

$$\mathcal{U}^n = \prod_{j=1}^{2^n-(n+1)} \mathcal{R}(S_j).$$

In Newton's method, an initial point determines a sequence $g_f(x_1)$. In voting, a given profile $\mathbf{p}$ determines the election ranking for each subset of candidates. Denote this listing, or *word*, by $F(\mathbf{p})$. For instance, if c_1 = milk, c_2 = beer, c_3 = wine and $\mathbf{p}$ is the profile of the beverage example, then $\mathbf{p}$ defines the word

$$F(\mathbf{p}) = (c_2 \succ c_1, c_3 \succ c_1, c_3 \succ c_2, c_1 \succ c_2 \succ c_3).$$

The complexity of Newton's method is analyzed by comparing the size of its dictionary with that of the universal set. Similarly, in voting, one way to understand "everything that can occur" is to characterize the *dictionary of election outcomes*

(3.1) $\mathcal{D}^n = \{F(\mathbf{p}) \in \mathcal{U}^n \mid \mathbf{p}$ is a profile, no restriction on the number of voters$\}$.

By introducing the dictionary, which consists of all possible words (listings of sincere election outcomes), the emphasis switches from local characteristics to a search for global properties. In particular, it now makes sense to search for everything that possibly could happen: i.e., to find all words in the dictionary. If this could be done, the entries of the dictionary would specify *all possible relationships and paradoxes of plurality election rankings.*

4. Comparing Dictionaries

Mimicking the approach used with Newton's method, the first goal is to determine whether $\mathcal{D}^n$ is a large or small subset of $\mathcal{U}^n$. Clearly, $\mathcal{D}^n$ contains all of those well-behaved words where the election ranking of each subset behaves as expected because they agree with one another; e.g., $(c_1 \succ c_2, c_1 \succ c_3, c_2 \succ c_3, c_1 \succ c_2 \succ c_3) \in \mathcal{D}^3$. As all remaining words lack this orderly property, they identify "paradoxes." Consequently, a dictionary with only a few words represents an orderly situation. Conversely, a large dictionary, with its rich selection of words specifying different ways election outcomes can vary over the subsets of candidates, suggests a chaotic state of affairs admitting many new election paradoxes. So, is $\mathcal{D}^n$ is a large or small subset of $\mathcal{U}^n$?

Theorem ([S1]). *For the plurality voting system, where each voter votes for his or her top-ranked candidate, and for any $n \geq 3$, we have that*

$$\mathcal{D}^n = \mathcal{U}^n. \tag{4.1}$$

This disturbing conclusion allows *any* sequence of rankings, one ranking chosen for each subset of candidates, to be admissible election outcomes! As such, it asserts that there is a profile supporting the flipping election outcomes $c_1 \succ c_2$, $c_3 \succ c_2 \succ c_1$, $c_1 \succ c_2 \succ c_3 \succ c_4$, $c_5 \succ c_4 \succ c_3 \succ c_2 \succ c_1, \dots$ where one kind of ranking applies with an even number of candidates but the reverse holds with an odd number. In fact, we can extend this example to require all subsets with an even number of candidates to be ranked consistent with $c_1 \succ c_2 \succ \cdots \succ c_n$, and all other subsets to be ranked in the opposite manner — or, in any other perverse way we may wish.

As another illustration, this theorem allows us to extend the beverage example to the extreme setting where the group's ranking is $c_1 \succ c_2 \succ c_3 \succ \cdots \succ c_n$, yet their ranking of *all proper subsets of candidates* is the exact reversal! To help the reader relate to this example, suppose this occurs when your department ranks n candidates for a single tenure track position. The first vote suggests that c_1 is the favorite, but, is she? Should *any* candidate withdraw, the department's sincere ranking is completely reversed! Who should get the offer: c_1 or c_n? Even worse, as we rarely hold another election just because some candidates withdraw, we might never discover that the reason we made a bad selection is due to this behavior.

Extension of the literature. Because the traditional literature tends to be highly property- and example-specific, it is clear how this theorem extends what was previously known. (For a review, I recommend the nice survey by Niemi and Riker [NR] and the books [Br, O, K2, Nu, St].) Namely, the traditional emphasis on local properties forces much of the literature to emphasize special cases and examples of the Fishburn type. Recall, his reversal example proves there exists a profile defining the contradictory plurality rankings $A \succ B \succ C \succ D$ and $C \succ B \succ A$. Clearly, there is no way to guess what else can accompany this behavior; e.g., it does not address what might happen should C, or B, or A, rather than D, withdraw from the competition. The above theorem, however, provides the answer: it asserts that the subsequent ranking can be whatever you want it to be! Just one possibility has *the ranking reversed whenever any candidate (not just the bottom-ranked one) withdraws.* In fact, the theorem allows you to extend this example in any imaginable way by using all subsets of candidates including the pairs. Similarly, all other published examples demonstrating any perverse outcome caused by candidates joining or withdrawing from the election can be significantly extended in previously unimaginable ways.

As another illustration, a troubling behavior known for two centuries and exploited in interesting, unexpected ways by contemporary authors including Arrow [A], McKelevey [M], and Kramer [Kr], is where the sincere pairwise elections define cycles such as $c_1 \succ c_2, c_2 \succ c_3, \dots, c_{n-1} \succ c_n, c_n \succ c_1$. (As shown in [S7], cycles and their supporting profiles can be understood with the orbits of a cyclic group of order n.) To the best of my knowledge, cycles involving larger subsets of candidates have not been discovered, and for good reasons — first, why should anyone even suspect that such perverse behavior exists, and, even with suspicions, it is difficult to find supporting profiles. It follows trivially from the theorem, however, that such phenomena exist (and are robust). The theorem requires, for instance, that there exist voters' preferences defining the plurality cycle $c_1 \succ c_2 \succ c_3$, $c_2 \succ c_3 \succ c_4$, $c_3 \succ c_4 \succ c_1$, $c_4 \succ c_1 \succ c_2$. In fact, we can further complicate this example by requiring these voters' election ranking for the pairs to cycle in the "other" direction $c_2 \succ c_1$, $c_1 \succ c_n, \dots$. Again, all of this holds because whatever rankings are specified for the different subsets, the theorem ensures that the plurality method allows them to coexist!

Another theme receiving considerable attention (e.g., see [NR, Nu] and the references they list) is to understand procedures, such as runoff elections or tournaments, where after dropping certain candidates another plurality vote is taken. But, is it possible for a candidate preferred by the voters to be dropped at the first step? How about a candidate who always is favored when compared with any other candidate: could she lose with a runoff election? Thanks to the theorem, we now can answer all questions of this type about almost all procedures. To illustrate with elimination methods, just rank the initial set of candidates so that c_1 is dropped. Then, for all other subsets containing c_1, have her top-ranked. Such examples, where c_1 is arguably the voters' favorite even though she is dropped at the first stage, exist because the theorem admits *any listing of rankings* for the subsets of candidates. So, while the literature can answer certain specific issues, this theorem motivated by dynamics provides a far more general, yet almost a trivial way to resolve a much wider spectrum of questions.

A related theme is to determine whether the choice of a procedures can influence the outcomes. (Again, see [Br, K, NR, Nu, O, St].) For instance, to choose who to hire for the one departmental opening we could use the plurality ranking, or, alternatively, we could plurality rank the candidates in subsets of five and then plurality rank the resulting set of winners. But, just by observing that these procedures involve different subsets of candidates, it is immediate from the theorem that their outcomes can be as different as desired. Proving this conclusion only involves assigning appropriate rankings to specified subsets of candidates. Once done, we still can obtain all sorts of other results just by recognizing that there are many other subsets of candidates where we haven't assigned a ranking. So, by choosing appropriate rankings for these sets, it is easy to prove how sensitive the outcome is to how the candidates are assigned to the five-candidate subsets, or to whether three, or four, or six, ... candidate subsets are used, or whether we use procedures involving pairwise votes, or whether a runoff election is used, or ... What a chaotic state of affairs! General results of this new type, of course, are impossible with the traditional local analysis.

As a final troublesome corollary, note that we can choose the rankings for each of the $2^n - (n+1)$ subsets in a completely random fashion. Even though these election rankings need not have anything to do with one another, the theorem ensures there is a profile where the sincere election outcome for each subset of candidates is the randomly selected one! This conclusion, which again demonstrates the power of this global perspective, doesn't instill much confidence in our standard tool of democracy.

Other election methods. Why use the plurality vote? Instead of just voting for our top-ranked candidates, maybe we should recognize voters' lesser ranked candidates. This approach was pioneered by the French mathematician J.C. Borda in 1770. He argued that with k candidates, a voter's top-ranked candidate should receive $k-1$ points, ..., the jth ranked candidate should receive $k-j$ points, ..., $j = 1, 2, \ldots, k$. More generally, the *Borda Count* (BC) is where the difference between points assigned to successive candidates is the same positive scalar. (So, for $n = 4$ candidates, both $(3, 2, 1, 0)$ and $(25, 20, 15, 10)$ define Borda Counts.)

Why this particular choice? Why not some other arrangement? After all, any vector $\mathbf{w} = (w_1, w_2, \ldots, w_k = 0)$ suffices as long as $w_j \geq w_{j+1}$ for $j = 1, \ldots, k-1$, and $w_1 > 0$. With such a *voting vector*, w_j points are assigned to a voter's jth ranked candidate and a candidate's election ranking is based on how many points she receives. The responses of Borda, Laplace, and other mathematicians from the late eighteenth century supporting the BC are more philosophical than mathematical so they are unsatisfactory. Nevertheless, the BC was used for years in the French Academy until changed by a new member Napoleon Bonaparte.

To investigate these questions, for each subset of candidates S_j, assign a voting vector $\mathbf{w}_j$. This defines a *system voting vector*

$$\mathbf{W}^n = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2^n-(n+1)}).$$

Thus, with an assigned system voting vector and a profile, the ranking of the jth subset of candidates S_j is found by tallying the ballots with $\mathbf{w}_j$. (Think of $\mathbf{w}_j$ as defining the "dynamic" on the S_j portion of the product space.) Let $F(\mathbf{p}, \mathbf{W}^n)$ be the resulting listing of these election rankings: $F(\mathbf{p}, \mathbf{W}^n)$ is the $\mathbf{W}^n$ word defined by $\mathbf{p}$. By admitting all possible profiles, we obtain the $\mathbf{W}^n$ dictionary

$$\mathcal{D}^n(\mathbf{W}^n) = \{F(\mathbf{p}, \mathbf{W}^n) \mid \mathbf{p} \text{ is a profile}\}.$$

Again, to understand the consequences of different system voting vectors, use the dictionary measure. Because system voting vectors can be identified with points in an appropriate dimensional Euclidean space, we use these mathematical structures to describe sets of system vectors. Let $\mathbf{B}^n$ be the system vector where the BC is used with each subset of candidates.

Theorem ([S1,S2]). *For $n \geq 3$, with the exception of a lower dimensional algebraic set α^n, all system voting vectors have the property that*

$$\mathcal{D}^n(\mathbf{W}^n) = \mathcal{U}^n.$$

If $\mathbf{W}^n \neq \mathbf{B}^n$, then

$$\mathcal{D}^n(\mathbf{B}^n) \subsetneq \mathcal{D}^n(\mathbf{W}^n). \tag{4.2}$$

In general, then, the election outcomes can be as chaotic as desired! Only those system voting vectors belonging to a particular algebraic set (which recently has been characterized [S3,4] and used to define a partial ordering for system vectors) can avoid certain paradoxes. Also, because Eq. 4.2 identifies the BC as the unique method to minimize both the number and kinds of paradoxes that can occur (for any $n \geq 3$), we finally have the sought after mathematical justification supporting the optimal status of the BC. With the beverage paradox, for instance, the BC ranking is the more reasonable "wine $\succ$ beer $\succ$ milk" with the tally 19 : 14 : 12. Indeed, the BC is the *only* voting method where its outcomes must be related to the pairwise rankings. For any other method, choose the rankings of the $\binom{n}{2}$ pairs and the set of all n candidates in any desired manner and there is a supporting profile; only the BC imposes order upon these electoral conclusions. ([S2, S7])

But, does this difference in dictionaries matter? If the BC avoids only a couple of paradoxes, then who cares? However, as dramatically illustrated with the inequality $10^{50}|\mathcal{D}(\mathbf{B}^6)| < |\mathcal{D}(\text{ Plurality}^6)|$, already with just 6 candidates the Borda Dictionary allows shockingly fewer paradoxes than, say, the plurality method. Many other arguments along this line, along with the characterization of the entries of $\mathcal{D}(\mathbf{B}^n)$ ([S2]), can be advanced to prove that the BC is, by far, the superior choice. (See, for example, [S3, S7].) This includes the single transferable vote used by the AMS, Approval Voting by the MAA, and plurality voting by SIAM.

Extending the literature. A natural mathematical theme is to discover invariants — here the invariants are relationships among election rankings. Of the comparatively few that were previously known is one asserting that the top-ranked BC candidate never could be beaten in all possible pairwise elections. (Most surely, Borda knew this result; it definitely was understood by Nanson [N] in the nineteenth century and then rediscovered by several others (including me) in recent years starting with Smith's nice paper [Sm].) As already indicated, this perverse phenomenon does occur with all other positional procedures!

An "election relationship," then, is the complement of the dictionary; it can be thought of as specifying what "paradoxes" cannot occur. Therefore, these ranking relationships are completely determined by the set $\mathcal{U}^n \backslash \mathcal{D}^n(\mathbf{W}^n)$. So, now that we know ([S3, 4]) how to find the entries of all dictionaries $\mathcal{D}^n(\mathbf{W}^n)$, *we also know all possible ranking relationships*. (Again, just as in dynamics, this is a consequence of changing from a local to a global perspective.) And, just by the number count given above for $n = 6$ candidates, it is clear that this dictionary approach uncovers an incredible number of new relationships previously not even suspected. One new Borda relationship, for instance, is that if there is an integer k, $2 \leq k < n$, so that c_1 is top-ranked in all k-candidate elections, then she is not bottom-ranked in the full election. Other new kinds of Borda relationships impose restrictions on the possible k-candidate rankings for each $k > 2$, etc., etc.

Further insight into how this theorem extends the literature exploits the assertion that, in general, "anything can happen." Consequently the earlier comments about runoffs, comparing procedures, etc. immediately extend from the plurality vote to almost all other positional voting methods. Again, most results (i.e., all that I know about) using the traditional local emphasis are able to compare only limited classes of procedures while addressing highly restricted questions. With this global approach (as true for dynamical systems), these conclusions can be extended in almost all ways with minimal effort.

To further illustrate the advantages of this global approach, I point to the attention focussed on the BC. (See almost any article listed in [K] with the word

"Borda" in the title.) In part, this is because the BC is well known and easy to use. Usually BC faults are identified by discovering specific profiles. (This means they have discovered portions of particular words from $\mathcal{D}^n(\mathbf{B}^n)$.) The obvious flaw with this approach is that we don't know whether the conclusion is specific to the BC, or whether the identical fault holds for other procedures. Equation 4.2, however, gives the answer. *Any BC fault illustrating undesirable changes in how the candidates are sincerely ranked over different subsets must be shared by all other procedures.* Namely, a BC flaw is universal: it must be suffered by all procedures. On the other hand, the strict containment of Eq. 4.2 means that all other procedures admit faults that are impossible with the BC.

The choice of a procedure matters. Before examining topics other than voting, one might wonder whether the choice of a voting method matters. For a given profile, won't all procedures give essentially the same outcome? As the beverage example already proves (the BC reverses the plurality outcome), the answer is no. (In fact, this profile admits *seven* different rankings as the weights change [S7].) The first general result in this direction seems to be where Fishburn [F1] proved that two different tallying methods can admit opposite outcomes. But now, by use of this more general perspective, we can find all possible rankings that occur by changing the procedure ([S4, S7]). As a dramatic example showing how this extends the literature, we now know [S4] that instead of just two reversed outcomes, it is possible for a ten-candidate profile to admit *over 84 million different rankings of the 10 candidates!* Remember, the voters' preferences remain fixed as marked on the ballots, so these millions upon millions of outcomes are due to changes in the tallying method. In fact, each of the ten candidates can be top-ranked when some procedures are used and then bottom-ranked with others! So, which of these highly contradictory but "sincere" rankings is the correct one? No wonder we need a mathematical theory to understand voting procedures!

Statistics. To illustrate this dictionary theme by generalizing Simpson's paradox from statistics, start with 2^j pairs of urns marked I_k^1, II_k^1. Place red and blue balls in each urn; the choice determines whether $P(R|I_k^1)$, the probability of choosing red when selecting at random from urn I_k^1, is larger, smaller, or equal to $P(R|II_k^1)$. Thus, associated with each pair are the three symbols $\{>, =, <\}$. In the empty urn I_k^2, combine the contents of I_{2k-1}^1 with I_{2k}^1; similarly, let II_k^2 hold the contents of II_{2k-1}^1 and II_{2k}^1, $k = 1, \ldots, 2^{j-1}$. Again, for each pair of urns, one of the three symbols applies.

Continue this aggregation process where, at the sth stage, the contents of I_{2k-1}^s and I_{2k}^s are combined to create I_k^{s+1}, while the contents of II_{2k-1}^s and II_{2k}^s are combined to create II_k^{s+1}, $s = 1, \ldots, j$, $k = 1, \ldots, 2^{j-s}$. For each of the $2^{j+1} - 1 = \sum_{i=0}^{j} 2^i$ pairs of urns, one of the symbols $\{>, =, <\}$ applies. Therefore, the universal set for the aggregation of urns, $\mathcal{U}_{agg}^j$, consists of $3^{2^{j+1}-1}$ listings of symbols. An initial condition, $\mathbf{p}$, is the initial allocation of red and blue balls in the first layer of urns, and the listing of inequalities for the different urns defined by $\mathbf{p}$ is a word. The dictionary for the aggregation of urns, $\mathcal{D}_{agg}^j$ consists of all words. Again, the complexity of the aggregation process is measured by the following.

Theorem. *For all $j \geq 2$ with the aggregation of urns,*

$$\mathcal{D}^j_{agg} = \mathcal{U}^j_{agg}.$$

Again, anything can happen. For instance, start with $2^{50} \approx 1.125 \times 10^{15}$ urns, and keep combining the contents in the above manner until only one pair is left. The initial contents can be chosen so that $P(R|I^s_k) > P(R|II^s_k)$ holds for each pair during the first 49 levels, but there is a reversal at the last stage. Figure 4 (designed by A. Konchan as a homework exercise) illustrates a $j = 2$ example where the inequalities flip at each level.

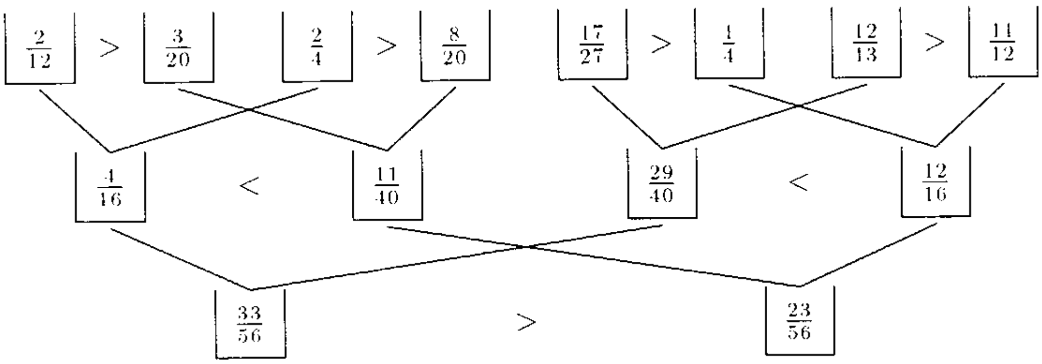

Figure 4. A three level Simpson paradox

Another illustration of the power of this dictionary approach comes from the widely used *Kruskal-Wallis* test of nonparametric statistics. This natural method replaces data with rankings. For instance, in comparing firms producing light bulbs according to the hundreds of hours a bulb lasted, the outcome

Firm 1	**Firm 2**	**Firm 3**
6.01	5.90	5.85
6.10	6.05	6.15

defines the KW matrix of ranks

	Firm 1	**Firm 2**	**Firm 3**
	3	2	1
	5	4	6
Total	8	6	7

and the KW ranking of Firm 1 $\succ$ Firm 3 $\succ$ Firm 2.

If bottom-ranked Firm 2 goes out of business, we want to compare Firms 1 and 3. The data remains fixed, but the relative ranking does not; now the KW ranking is Firm 1 $\sim$ Firm 3. While this is a reasonably innocuous change, we might wonder whether more radical outcomes can occur. This question was answered by Deanna Haunsperger in her thesis [H1] (part of which is in [H2]).

With $n \geq 3$ firms, there are $2^n - (n+1)$ subsets of two or more firms. Thus, the subsets and the symbols are the same as those for the voting problem. In place of voters' preferences, the initial data represents the initial condition, and the list of rankings for each subset of alternatives defines the KW word. The KW dictionary, $\mathcal{D}^n_{KW}$, consists of all possible words admitted by the KW test. In characterizing $\mathcal{D}^n_{KW}$, Haunsperger proved that the Kruskal-Wallis test admits far more paradoxes and difficulties than previously imagined. For instance, for any $n \geq 3$, examples of data can be found so that even though the KW rankings of pairs is Firm i $\succ$ Firm j iff $i < j$, the KW ranking comes perilously close to reversing this conclusion as it is Firm $n-1 \succ$ Firm $n-2 \succ \ldots \succ$ Firm $1 \succ$ Firm n. (The argument in [S7] discussing pairwise voting rankings extends to this statistical setting to show that we should place trust in the full KW ranking – the real difficulty is caused by deficiencies in the pairwise rankings.)

On the other hand, even though the KW test admits previously unknown difficulties, Haunsperger also proved that of all possible nonparametric ranking methods, the KW test is by far the best choice. She did this by showing that the dictionary for most nonparametric methods agrees with the universal set. Then she proved that the KW dictionary is a proper subset of the dictionary for any other method. In addition, she proved that in certain settings, the KW dictionary agrees with the Borda dictionary; the "symmetry" reasons for this are suggested below.

Probability. The cyclic dice example can be generalized to all levels as follows. (See [FS, Ho].) For $j \geq 2$, start with 3^j triplets of dice, $< A^1_k, B^1_k, C^1_k >$. There is a constructive approach to mark these dice so that $A^1_k \succ B^1_k, B^1_k \succ C^1_k$, and $C^1_k \succ A^1_k$ for all k. Now, treat each triplet as a set of dice; e.g., denote by A^2_k the three dice in the triplet $3k-2$ (so $A^2_k =< A^1_{3k-2}, B^1_{3k-2}, C^1_{3k-2} >$), B^2_k is the triplet $3k-1$ and C^2_k is the triplet $3k$, $k = 1, \ldots, 3^{j-1}$. Instead of "high score wins," with the new triplets, the larger sum of the set of dice wins. Again, a cycle occurs. Indeed, continue the aggregation process; at each aggregation level the dice define the cycle

$$A^s_k \succ B^s_k, B^s_k \succ C^s_k, \text{ and } C^1_k \succ A^1_k; s = 1, \ldots, j; k = 1, \ldots, 3^{j-s+1}.$$

Each pair of dice admits the three possible symbols $\succ, \sim, \prec$. Thus, a triplet admits 3^3 possible symbols. The universal space for the dice, $\mathcal{U}^j_{dice}$, is the space of listings of these symbols for each of the $3^{j+1} - 1$ triplets. The initial condition corresponds to how the dice are marked. Each choice of markings defines a sequence of symbols – a word in the dictionary $\mathcal{D}^j_{dice}$. A measure of the complexity of these dice games, which significantly extends our earlier construction, follows.

Theorem. *For the dice problem*

$$\mathcal{D}^j_{dice} = \mathcal{U}^j_{dice}.$$

Again, anything can happen! Indeed, to show how wild the process can be, Funkenbusch created an example [Ho] showing how the final stack of dice can be split in many different ways allowing for all sorts of cyclic outcomes.

Complexity of aggregation. The point is made. Clearly, when this dictionary construction is applied to other aggregation and classification procedures, examples emerge demonstrating a similar complexity while exposing many new paradoxes. In fact, related assertions have been found even for processes involving function spaces and vector fields (rather than finite discrete objects); assertions which raise doubts about such commonly accepted themes as the "supply and demand" story and other techniques from economics [S5]. What these examples illustrate is that aggregation processes, the basis of statistics, probability, and much of the social sciences, are far more complex than previously expected.

To provide an overview of the rest of the story, I briefly outline how other structures from dynamical chaos can be identified with standard concerns from aggregation processes. For instance, it is natural to worry about the effects of small data errors for statistics and small numbers of voters trying to manipulate the outcome. While I haven't defined "manipulation," you know what I mean – it is what you did the last time you cleverly marked your ballot to try to get a better outcome.

Typically (at least for voting), these issues are studied by using computer simulations. (See, for instance, [C].) In our formulation, as they involve comparing changes in outcomes due to small changes in a profile, these concerns are identified with the "sensitivity with respect to initial conditions." This motivates the development of analytic techniques to analyze these issues. For instance, it is natural to try to determine which procedures are least susceptible to these "small change," negative effects. From this analytic approach we learn that the voting system least susceptible to small manipulation problems is the BC [S7]; for nonparametric systems, the answer is the KW test.

Another natural concern is to determine the likelihood of paradoxes. (While most of this analysis involves computer simulations, [G, GF] have some analytic voting results for $n = 3$.) The parallel theme from dynamics is where the "dictionary" and "topological" approaches of chaos have been refined by using measure theory and topological entropy to indicate the likelihood of various behaviors. (See, for instance, [ALM, R].) Again, using the approach motivated by dynamics, work by Van Newenhizen [VN] and [S7] significantly extend these earlier efforts and again prove that the BC is the least likely to admit different paradoxes. One must anticipate a similar conclusion for the KW statistical test.

As a final illustration, there is an enormous literature (e.g., see the expository books [Nu, St]) identifying perverse behavior such as where a candidate can be hurt by receiving more electoral support. In our formulation, this turns out to be identified with the Cantor set structure from dynamics. While Cantor sets don't occur for these aggregation procedures (because finite, not infinite intersections are involved), the geometry of data generated by the iterated inverse images can be complicated. In fact, most (i.e., all that I know of) of the identified difficulties plus many new ones can be explained when convexity, connectiveness, and other geometric properties are lost by this iterative process. In this way it becomes easy to discover and explain examples such as where a department splits into two subcommittees to select one of three candidates for Chair. With the standard runoff (where the top two candidates from a plurality election have a runoff) it could be that each subcommittee chooses c_1 only to have c_1 lose when the full department uses the same procedure. (For an example and a geometric explanation showing that this is a convexity issue, see [S7].) A similar result holds for other aggregation methods from statistics and elsewhere that involve the rankings from different subsets of the alternatives.

5. Symmetry

We can think of these theorems in terms of the iterated inverse image approach outlined for Newton's example. (This is how the results in [S6] are proved.) However, the associated analysis quickly becomes overly complicated. Again, help to develop alternative approaches comes from modifying ideas developed in dynamical systems.

In dynamical systems, the iterated inverse image approach rarely is used. Instead, other evidence (e.g., "horseshoes" or "period-three orbits") is employed to indicate whether inverse images are empty or not. (See, for example, [D, LY].) For instance, one way to develop techniques and understand where bifurcations occur is to determine whether the boundary points of the inverse image of a region are from turning points (critical points of the mapping) or boundary points of the region. In a conceptually similar manner, singularity theory is used to identify what occurs with aggregation problems. Replacing boundary points are the regions involving a "tie vote." "Turning points" correspond to certain singularity structures from the orbits of symmetry groups.

Symmetry is the key. The natural symmetry action in voting is the permutation group $\mathcal{S}_n$ specifying the $n!$ "voter types" (i.e., the ways voters can rank the candidates). So, if $A^n = c_1 \succ c_2 \succ \cdots \succ c_n$, then any other voter type is obtained with an appropriate permutation $\sigma(A^n)$, $\sigma \in \mathcal{S}_n$.

Assume $\mathbf{w}^n$ is assigned to tally the ballots for the n candidates. Because $\mathbf{w}^n$ and $\lambda\mathbf{w}^n$, $\lambda > 0$, always define the same election ranking, assume that $\mathbf{w}^n$ has unit length. Next, if the jth component of a vector from R^n is identified with the tally for c_j; then $\mathbf{w}^n$ represents the *vector tally or ballot* for a voter with the preference A^n. The vector ballot for a voter with type $\sigma(A^n)$ is a permutation of $\mathbf{w}^n$ represented by $\mathbf{w}^n_\sigma$. In this manner, the $\mathcal{S}_n$ orbit of $\mathbf{w}^n$ defines all possible vector ballots. The rational points in the convex hull of $\{\mathbf{w}^n_\sigma\}_{\sigma\in\mathcal{S}_n}$ identify all possible election outcomes. This is because a point in the hull has a representation $\sum_{\sigma\in\mathcal{S}_n} t_\sigma \mathbf{w}^n_\sigma$ where $t_\sigma \geq 0$, $\sum_{\sigma\in\mathcal{S}_n} t_\sigma = 1$. So, rational values of t_σ can be identified with the fraction of all voters with the voter type $\sigma(A^n)$.

The appropriate structure to compare election outcomes over all $2^n - (n+1)$ subsets of two or more candidates is the orbit of the system vector $\mathbf{W}^n$. Here, for a voter with preferences A^n, each component of $\mathbf{W}^n$ identifies the combined vector ballot over the subsets of candidates (in the obvious product space). For a voter of type $\sigma(A^n)$, designate the vector ballot by $\mathbf{W}^n_\sigma$. As above, the convex hull $\{\mathbf{W}^n_\sigma\}_{\sigma\in\mathcal{S}_n}$ identifies all possible election outcomes. Therefore, all of the election properties of $\mathbf{W}^n$; including those described in terms of the dictionaries as well as many others, must be consequences of the orbit $\{\mathbf{W}^n_\sigma\}$. In particular, as the dimension of this hull increases, so do the kinds of admissible election outcomes.

To understand $\{\mathbf{W}^n_\sigma\}$, the first step is to recognize that rather than being the orbit of $\mathcal{S}_n$, $\{\mathbf{W}^n_\sigma\}$ is the orbit of a more complicated group action. To see why with $n = 3$, observe that when $\sigma = (1,2)$ is applied to the subset $\{c_2, c_3\}$, the outcome can vary. If σ acts on $c_1 \succ c_2 \succ c_3$, then σ has no impact upon the ranking of the subset. But when σ is applied to $c_1 \succ c_3 \succ c_2$, σ reverses the ranking of the subset. Thus the effect of a permutation from $\mathcal{S}_n$ depends not only on the choice of a permutation, but also on the element of the orbit to which it is applied. This defines a rich group structure known as the wreath product (of the permutation groups for each subset of candidates.) Related wreath product structures are needed for the different examples from statistics, probability, etc. given above.

Following the lead of dynamics and group theory, the natural question is to determine the singularities of this group action: namely, is there a $\mathbf{W}^n$ where (the convex hull) of its orbit has the smallest dimension? This occurs when the BC is assigned to each subset of candidates. What causes the singularity is the symmetry requirement that the difference between successive weights agree: a similar symmetry argument applies for the KW weights.

From this observation, several results are forthcoming. First, since the "dimension" of the hull defined by an orbit corresponds to the number and kinds of paradoxes that can occur, we can (correctly) expect the BC to minimize what "can go wrong." ([S1-4, 7]) Moreover, just as when latitude lines on the sphere (the $SO(2)$ orbits) decrease in size as they approach the singular orbit defined by the North Pole, as $\mathbf{W}^n$ approaches the BC with its singular orbit, the likelihood of these electoral difficulties also decreases. (For $n = 3$, some of these properties have been extracted and developed in [S7] with elementary geometric arguments.) Thus, for instance, we must (accurately) expect it to be less likely for paradoxes to occur with $(2, \frac{1}{2}, 0)$ than with $(1, 0, 0)$, or with $(3, 2, 0, 0)$ than with $(1, 0, 0, 0)$. In both cases, when normalized, the first vector is closer than the second to the BC vector.

Singularity theory tells us that the singular orbits form a stratified structure. This stratification corresponds to an algebraic set of procedures (e.g., system voting vectors, statistical ranking methods, etc.) which admit increasing levels of consistency in outcomes. Moreover, the stratification relates the outcomes admitted by procedures from different stratified levels, so it defines a partial ordering among the system voting vectors; part of this ordering is given in Eq. 4.1 showing that the Borda Dictionary is a proper subset of all other dictionaries. More generally, this structure allows a characterization of all possible (ranking) properties of each procedure. (See [S3, 4, 7].)

To suggest other properties disclosed by the symmetry, notice that by being lower dimensional, a singular orbit satisfies other symmetry relationships. Thus, the BC should satisfy properties denied to other procedures. For instance, the $\mathcal{S}_n$ symmetry of positional voting methods, called "neutrality," means that if all voters interchange the names of the candidates in a manner defined by a specified $\sigma \in \mathcal{S}_n$, then the outcome also changes in this manner. Similarly should all voters completely reverse their ranking of the candidates; the election ranking also should be completely reversed. But, this is false! For $n = 3$ candidates, only the BC respects this reversal symmetry. Instead, all other positional methods even permit the reversed profile to return the same normalized election tally! (See [S7].) This identifies a new class of paradoxes suffered by methods other than the BC.

What happens is that the $n!$ ways to rank the candidates define two permutation groups $\mathcal{S}_n$ and $\mathcal{S}_{n!}$. A name change of the candidates $\sigma \in \mathcal{S}_n$ defines a permutation, $\sigma_T \in \mathcal{S}_{n!}$, on the profile space. Neutrality requires $f(\sigma_T(\mathbf{p})) = \sigma(f(\mathbf{p}))$; i.e., it requires f to commute with the *neutrality subgroup* $\mathcal{N}_n = \{\sigma_T\}_{\sigma \in \mathcal{S}_n}$. However, the permutation from $S_{n!}$ defined by reversing each voter's ranking is not in $\mathcal{N}_n$. The BC, with its singular orbit, does preserve this reversal property; i.e., the BC commutes with a larger subgroup $\mathcal{BC}_n$ of $\mathcal{S}_{n!}$. Similar statements hold for other aggregation procedures showing that the strength and desired properties of the BC, KW, and other symmetric aggregation procedures derive from the singularity structure of their orbits.

REFERENCES

[A] Arrow KJ, *Social choice and individual values, 2nd edn.*, Wiley, New York, 1963.

[ALM] Alseda, L., J. Llibre, M. Misiurewicz, *Combinatorial Dynamics and Entropy in Dimension One*, World Scientific, 1993.

[BO] Bartholdi, JJ III, J. A. Olin, *Single transferable vote resists strategic voting*, Social Choice Welfare **8** (1991), 341-354.

[BTT] Bartholdi, JJ III, C. A. Tovey, and M. A. Trick, *How hard is it to control an election?*, Mathl. Comput. Modeling **16** (1992), 27-40.

[B] Blyth, C, *On Simpson's paradox and the Sure-Thing Principle*, JASA **67** (1972), 364-366.

[Br] Brams, S, *Paradoxes in Politics*, Free Press, New York, 1976.

[C] Chamberlin, JR, *An investigation into the relative manipulability of four voting systems*, Behav Sci **30** (1985), 195-203.

[D] Devaney, R., *An introduction to chaotic dynamical systems; 2nd ed.*, Addison-Wesley, 1989.

[F1] Fishburn, PC, *Inverted orders for monotone scoring rules*, Discrete Appl Math **3** (1981), 27-36.

[FS] Funkenbusch, B, and DG Saari, *Preferences among preferences or nested cyclic stochastic inequalities*, Congressus Numerantium **39** (1983), 419-432.

[G] Gehrlein, WV, *Expected probability of Condorcet's paradox*, Econ Lett **7** (1981), 33-37.

[GF] Gehrlein, WV, Fishburn, PC, *Probabilities of election outcomes for large electorates*, J Econ Theory **19** (1978), 38-49.

[GM] Good, I.J., and Y. Mittal, *The amalgamation and geometry of two-by-two contingency tables*, The Annals of Statistics **15** (1987), 694-711.

[Ho] Honsberger, R., *More mathematical morsels*, MAA, 1991.

[H1] Haunsperger, D., *Projection and aggregation paradoxes in nonparametric statistical tests*, Northwestern University PhD Dissertation, 1991.

[H2] Haunsperger, D., *Dictionaries of paradoxes for statistical tests on k samples*, JASA **87** (1992), 149-155.

[HS] Haunsperger, D. and D. G. Saari, *The lack of consistency for statistical decision procedures*, The American Statistician **45** (1991), 252-255.

[K] Kelly, JS, *Social choice bibliography*, Soc Choice Welfare **8** (1991), 97-169.

[K2] Kelly, JS, *Social Choice Theory*, Springer Verlag, New York, 1988.

[Kr] Kramer, G.H., *A dynamical model of political equilibrium*, J. Econ. Theory **12** (1977), 310-344.

[LY] Li T-Y and Yorke, J.A., *Period three implies chaos*, Amer Math Monthly **82** (1975), 985-992.

[M] McKelevey, RD, *Intrasitivities in mulitdimensional voting models and some implications for agenda control*, J. Econ Theory **12** (1976), 472-482.

[N] Nanson, EJ, *Methods of election*, Trans Proc R Soc Victoria **18** (1882), 197-240.

[NR] Niemi, RG, WH Riker, *The choice of voting systems*, Scientific American **234** (1976), 21-27.

[Nu] Nurmi, H., *Comparing Voting Systems*, D. Reidel, Dordrecht, 1987.

[O] Ordeshook, P., *Game Theory and Political Theory*, Cambridge University Press, 1986.

[R] Robinson, C., *Dynamical Systems*, Northwestern University book manuscript, 1994.

[S1] Saari, D. G.,, *A dictionary for voting paradoxes*, Jour Econ Theory **48** (1989), 443-475.

[S2] Saari, D.G., *The Borda Dictionary*, Soc Choice Welfare **7** (1990), 279-317.

[S3] Saari, D. G., *Symmetry extensions of "neutrality" I. Advantage to the Condorcet loser*, Soc Choice Welfare **9** (1992), 307-336.

[S4] Saari, D. G., *Symmetry extensions of "neutrality" II. Partial ordering of dictionaries*, Soc Choice Welfare **10** (1993), 301-334.

[S4] Saari, D.G., *Millions of election rankings from a single profile*, Soc Choice Welfare **9** (1992), 277-306.

[S5] Saari, D. G., *The aggregate excess demand function and other aggregation procedures*, Economic Theory **2** (1992), 359-388.

[S6] Saari, D.G., *The ultimate of chaos resulting from weighted voting systems*, Advances in Applied Math **5** (1984), 286-308.

[S7] Saari, D. G., *Geometry of Voting*, Springer-Verlag, New York, 1994.

[SU] Saari, D. G., and Urenko, J., *Newton's method, circle maps, and chaos*, Amer. Math Monthly **91** (1984), 3-17.

[Sm] Smith, JH, *Aggregation of preferences with variable electorate*, Econometrica **41** (1973), 1027-1041.

[St] Straffin, P., *Topics in the Theory of Voting*, Birhauser, Boston, 1980.

[VN] Van Newenhizen, J., *The Borda Count is most likely to respect the Condorcet principle*, Economic Theory **2** (1992), 69-83.

Department of Mathematics, Northwestern University, Evanston, Illinois 60208-2730